A Heterogeneous Web Service Integration Method Based-on Semantic Matching

Jie He[1], Nengcheng Chen[2], Wenbao Mi[1]

1 Ningxia University, Ningxia, China

2 Wuhan University, Wuhan, China

Email: whuhejie@yahoo.com.cn

KEYWORDS: version negotiation, semantic schema matching, information retrieval, data fusion

In order to achieve the spatial information sharing and interoperability, there are many international organizations such as the OGC, ISO/TC211, FGDC are working to develop a series of standards and norms, such as Web Service-based technology, OGC proposes seamless online geographic information processing services and location-based services framework (OWS), which provides a series of abstract specifications and implementation specifications, such as the Web Feature Service (WFS), Web Coverage Service (WCS), Sensor Observation Service specification (SOS) and for different types of web services sharing and interoperability in technology to a certain extent. Many international organizations and institutions have also begun the geospatial information interoperability substantive research. Project funded by the European Union (ECDG-Ⅲ) GIPSIE (GIS Interoperability Project Stimulating the Industry in Europe), presided over by the United States National Geographic Information and Analysis Center (NCGIA), the National Science Foundation-funded Varenius project, and the DISGIS project completed by the Norwegian Bureau of Surveying and Mapping units. And there are a lot of organizations have developed a variety of web services products, such as the GeoTools, Deegree, GeoServer and GeoSurf software, etc. Different products realize web service capabilities differ, not all type of web service, every version of a type of web service can achieve. So far, the interoperability between heterogeneous web services (i.e, homogeneous and heterogeneous web services) is still a problem for further study, such as different types of web services (heterogeneous network services) can't be achieved directly interoperable. Access to different versions of the same types of services (homogeneous web service) only by a version negotiation mechanism to achieve a mutually acceptable version of the service, but this mechanism has two major drawbacks: 1. can't meet the user accurate version of the service request; 2. Poor reliability, there is a potential risk of negotiation failure. According to the problems exist currently in integrate heterogeneous web services, resulting in a number of different solutions, such as Ching chien (2004) proposed a general geographic data integration framework to access and extract heterogeneous geographic data sources, through the use of dynamic synthesis operation to integrate geographic data. An ontology-driven geographic information systems is proposed by Frederico (2002), through the ontology to reduce the heterogeneity between the heterogeneous web services or data. A geographic ontology integration method proposed

by Nudelman (2006), while the design of geographic ontology matcher named G-Match to match the integrated geographic ontologies. Due to the definition of ontology elements concept in different granularity，the weighting factor calculated dependent on the input ontology itself. SWING (Semantic Web-Service Interoperability for Geospatial Decision Making) provides an open, easy-to-use semantic network services framework for ontology and reasoning tools to annotation, discovery, composition and invocation of web services. Thus ,current web service integration method rely on ontology matching, and the ontology is usually domain-oriented, lack of generic efficient ontology matching methods. Unfortunately, so far, the web service matching problem has not been solved by OAEI (OntologyAlignment Evaluation Initiative).

Based on the above background, this paper presents a heterogeneous web service integration method based on dynamic semantic schema matching and automatic information retrieval and fusion. The central idea including: 1) the exact schema matching. Using a semantic schema matching method based on node similarity [16] to achieve an exact match for heterogeneous web services schema; 2) automatic information retrieval and data fusion. Semantic schema matching results-based for ① homogeneous web services: schema matching mapping relationship (one-to-one) transformed into a different document information retrieval rules using XSLT and other technologies to achieve efficient conversion between different documents; ② heterogeneous web services: according to the mapping relationship between heterogeneous web service schema elements (one-to-one, one-to-many), to design heterogeneous web service fusion model, realize data been copied, composed, decomposed, converted between heterogeneous web services instance data. Eventually achieve the correct uniform access to heterogeneous web services and efficient integration of them.

The system includes three core components, respectively, is the schema matcher, the information extractor and data fuser. Schema matcher is designed to achieve exact match for heterogeneous web services model; information extractor is to achieve automatic conversion and integration of the different versions of homogeneity web service instance; data fuser is designed to realize the integration of heterogeneous web service instance data. For the same version of the same types of service requests can be directly dispatched to the server to implement a service request, the integrated access to the services of the different versions of the same type service including two steps, the one is schema matching between different versions of service schemas, and the other is information extraction and conversion between different documents. In the first stage, a series of internal models correspondence relationship specifications into a series of design choices mappings between two schemas. The design choices include hierarchical organization of data and schema constraints (such as foreign key constraints). In the second stage, these mappings translate into query on source schema by generating data filling to the destination schema (SQL, XQuery, XSLT). An important feature of the mapping algorithm is considering the destination schema constraint condition in order to ensure the generate data will not violate the integrity of the destination schema. For the heterogeneous web services, that is, the different types of services, in addition to the

schema translation request, before the service conversion would also like to fuse different information models, such as fusion between O&M result model and GML WCS profile or GML features summary table, and a real-time observation metadata generated from SOS registered in the CSW server into a “objectType”. If the observed results are coverage, the metadata “WCSCoverage” generated; if the observation results are features, “WFSLayer” metadata generated. “WCSCoverage” contains the name of the data set, the geographic range of data, formats, spatial reference, and resolution. “WFSLayer” contains the data set name, feature type, geographic range, spatial reference. Different types of data fusion operations are designed to integrate heterogeneous service instances intelligently as well.
The main contributions the study offers as following:

- Overcome the defects of version negotiation mechanism, achieves to access the exact version of web service.
- Apply fragment-based matching method to improve the matching efficiency, through the application of semantic matching method to improve the matching accuracy.
- Using service-oriented architecture, design a plug-and-play middleware, greatly improving system scalability and flexibility.

The rest of the article is organized as follows. The architecture, the main components of the system is described in section 2. System implementation is discussed in section 3. Section 4 discusses the experiments and discussion. Finally, section 5 summarizes the conclusions.